Revised by

Deema . Fatima . Talal . Ahmed . Mohammed
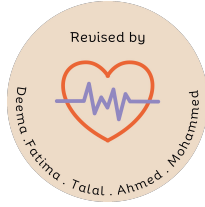
# Sampling techniques

Lecture No. 20

Objectives:

1.  Why should we use sampling methods ?
2.  To know definitions of few terms in sampling
3.  What are the different Sampling and non-sampling methods
4.  And you should able to use sampling methods appropriately in research process

~ This lecture was presented by **Dr. Leena Baghdadi,** and this **lecture** from **439**
~ It is included in the **Midterm Exam**
~ We highly recommended reading the **Ayah** in the first page

## Slides

**Color code**

Original text
Dr. Notes
Important
Golden note 🎁
Extra

## Editing file

## What is Sampling

Sampling is the process or technique of selecting a study sample of appropriate characteristics and of adequate size.

## Why to use Sampling in Research

- Unable to study all members of a population
- Reduce selection bias
- Feasibility, save time and money
- Measurements may be better in sample than in entire population

## Definitions (Q)

**Population:**
The group to which the results of the study will be generalized or applicable. A set which includes all measurements of interest to the researcher.

**Target population:**
the group from which the sample ultimately is selected. Who you will approach to collect your data.

**Study sample:**
the individuals who will volunteer or be selected from the target population and are the final source of data, and the final group from whom complete data will be collected. Must be **similar** to population on **characteristic** being investigated, and it's a subset of the population.

سُورَةُ الزُّخْرُفِ

وَلَوْلَا
أَن يَكُونَ ٱلنَّاسُ أُمَّةً وَٰحِدَةً لَّجَعَلْنَا لِمَن يَكْفُرُ بِٱلرَّحْمَٰنِ لِبُيُوتِهِمْ سُقُفًا مِّن فِضَّةٍ وَمَعَارِجَ عَلَيْهَا يَظْهَرُونَ ۝ وَلِبُيُوتِهِمْ أَبْوَٰبًا وَسُرُرًا عَلَيْهَا يَتَّكِئُونَ ۝ وَزُخْرُفًا ۚ وَإِن كُلُّ ذَٰلِكَ لَمَّا مَتَٰعُ ٱلْحَيَوٰةِ ٱلدُّنْيَا ۚ وَٱلْءَاخِرَةُ عِندَ رَبِّكَ لِلْمُتَّقِينَ ۝
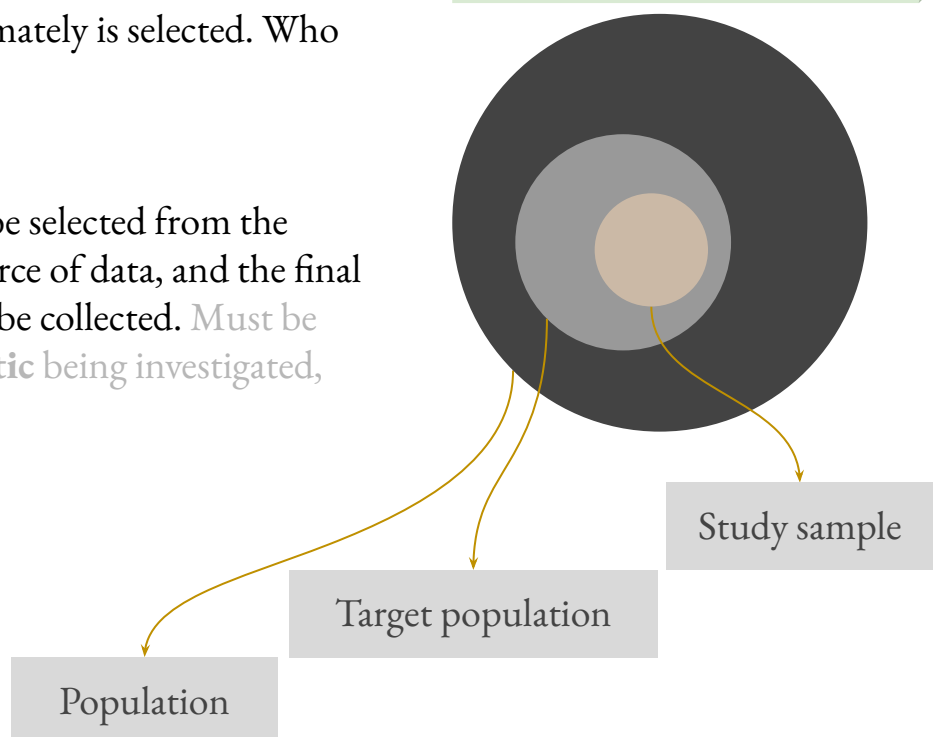
المختصر في التفسير

ولولا أن يكون الناس أمة واحدة في الكفر لجعلنا لبيوت من يكفر بالله سقوفًا من الفضة، وجعلنا لهم درجًا عليه يصعدون.
وجعلنا لبيوتهم أبوابًا، وجعلنا لهم أسرّة عليها يتكئون استدراجًا لهم وفتنة.
ولجعلنا لهم ذهبًا، وليس كل ذلك إلا متاع الحياة الدنيا، فنفعه قليل لعدم بقائه، وما في الآخرة من النعيم خير عند ربك - أيها الرسول - للمتقين لله بامتثال أوامره واجتناب نواهيه.

Population: the group I would like to generate the results on (most general)

Target population: a segment of the population, those I wait to approach to collect data (most specific)

Study population or sample: the final source of information (more specific)



Study sample

Target population

Population

| | |
|---|---|
| **Sampling Frame** | • This is the complete list of sampling units (study subjects) in the target population to be subjected to the sampling procedure.<br>• Completeness and accuracy of this list is essential for the success of the study.<br>• **Example (e.g.), Group B attendance sheet, it has to be complete with all names.** |
| **Sampling Units** | • These are the **individual units / entities** that make up the frame just as elements are entities that make up the population.<br>• **E.g., Each member from the sampling frame is sampling unit.** |
| **Sampling Error**<br>**For your own knowledge** | • This arises out of random sampling and is the **discrepancies** (inconsistency) between **sample values and** the **populations'** value. |
| **Sampling Variation**<br>**For your own knowledge** | • Due to infinite variations among individuals and their surrounding conditions.<br>• Produce differences among samples from the population and is **due to chance.**<br>• E.g., In a clinical trial of 200 patients we find that the efficacy of a particular drug is 75%. If we repeat the study using the same drug in another group of similar 200 patients we will **not get** the **same efficacy** of 75%. It could be 78% or 71%.<br>**"Different results from different trails, although all of them conducted under the same conditions"** |
| **Decisions Required for selecting sample** | 1. Specify what is the **target population.** This is entirely determined **by** the research **objective**.<br>2. Specify what is the **study population.** (e.g., who are **eligible for inclusion** in the study).<br>3. Select a **sampling design (technique)** for obtaining a sample for study.<br>4. Strategy to **ensure** high **response** or participation **rate**, otherwise inference must take account of non-responses.<br>Decisions will have considerable impact on study validity. |

## Representativeness (validity)

A sample should accurately reflect distribution of relevant variable in population

1.  Person (e.g., age and gender)
2.  Place (e.g., urban vs. rural )
3.  Time

Representativeness is essential to generalise the results.

Ensure representativeness before starting, and confirm once completed.

Study participants: those who agree to participate

**Validity of a Study has two components of validity:**

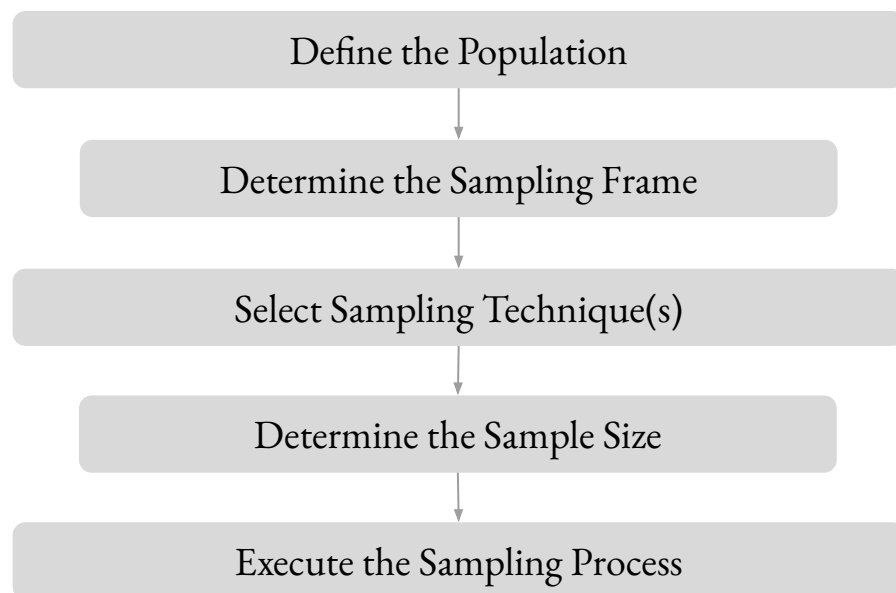| | 1- Internal validity | 2- External validity |
|---|---|---|
| **Description** | - A study is said to have internal validity when there have been proper selection of study group and a lack of error in measurement.<br>- It is Concerned with the appropriate measurement of exposure, outcome, and association between exposure and disease.<br>- You must have proper selection of the study group and their measurement to say that you have good internal validity. | - External validity implies the ability to **generalize** beyond a set of observations to some **universal** statement.<br><br>- For instance, do the findings apply to other people, settings, situations, and time periods? |
| **Study participants** | Study participants may still be not representative of the target population even with random sampling because of:<br>● **Sampling frame** is out of date<br>● Failure to **recruit** eligible subjects<br>● Non **consent** or non-**response**<br>● **Drop Out**/Withdrawal | |

## Sampling and representativeness

**Target Population → Sampling Population → Sample**

**How to sample:**

1. **Sampling frame** must be available, otherwise develop a sampling frame.
2. Choose an appropriate **sampling method** to draw a sample from the sampling frame.
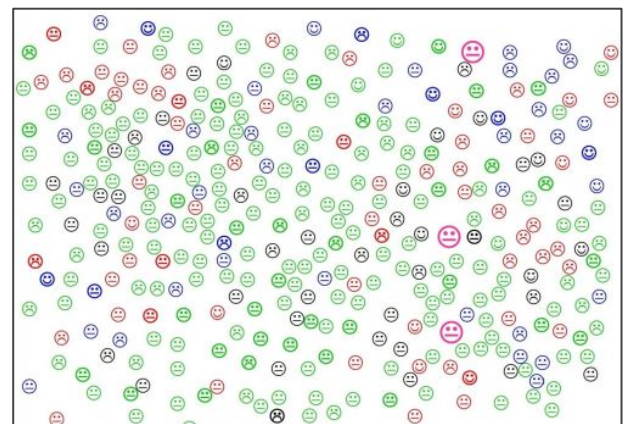
## The Sampling Design Process

Define the Population

↓

Determine the Sampling Frame

↓

Select Sampling Technique(s)

↓

Determine the Sample Size

↓

Execute the Sampling Process

## Who are They?

- Black / Blue / Green / Red
- Thin / Bold
- Smiling / Normal / Sad

Consider this picture is your target population, there are a lot of variety someone happy, someone sad etc.

# Types of Sampling Methods

**Probability Sampling**

- Simple Random Sampling (SRS)
- Stratified Random Sampling
- Systematic Random Sampling
- Cluster Sampling
- Multistage Sampling

**Non- probability Sampling**

- Deliberate (Quota) Sampling
- Convenience Sampling
- Purposive Sampling
- Snowball Sampling
- Consecutive Sampling

## 1. Simple Random Sampling

- **Equal probability,** crude type of data selection

- **Techniques:**
    - Lottery method
    - Random number table method

- **Advantage:**
    - Simple process and easy to understand
    - Easy calculation of means and variance
    - Most representative group, Avoids selection bias

- **Disadvantage:**
    - Not most efficient method, that is, not the most precise estimate for the cost
    - Cannot always be certain that there is an equal chance of selection
    - Non-respondents or refusal

## How to select a simple random sample?

1. Define the population
2. Determine the desired sample size
3. List all members of the population or the potential subjects

- **For example:**
  - A tattslotto draw is a good example of simple random sampling.
  - A Sample of 6 numbers is randomly generated from a population of 45, with each number, having an equal chance of being selected.

**Table of random numbers:**

- They are used after numbers have been assigned to numbers of the study population.
- Use random number table to select subject start anyway anywhere.
- Continuous selecting until the desired sample is reached.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 49486 | 93775 | 88744 | 80091 | 92732 |
| 94860 | 36746 | 04571 | 13150 | 65383 |
| 10169 | 95685 | 47585 | 53247 | 60900 |
| 12018 | 45351 | 15671 | 23026 | 55344 |
| 45611 | 71585 | 61478 | 87434 | 07498 |
| 89137 | 30984 | 18842 | 69619 | 53872 |
| 94541 | 12075 | 30771 | 19598 | 96069 |
| 89920 | 28843 | 87599 | 30181 | 26839 |
| 32472 | 32796 | 15255 | 39636 | 90819 |

### Potential Subject Pool

| | | |
|---|---|---|
| 1. Robert | 11. Ken | 21. Steve |
| 2. Ralph | 12. Wilmer | 22. Larry |
| 3. John | 13. Alan | 23. Rick |
| 4. Andy | 14. Kevin | 24. Bruce |
| 5. Joel | 15. James | 25. Clyde |
| 6. Thomas | 16. Henry | 26. Sam |
| 7. Cooper | 17. Don | 27. Kent |
| 8. Maurice | 18. Walt | 28. Travis |
| 9. Terry | 19. Doug | 29. Woody |
| 10. Carl | 20. George | 30. Brian |

So our selected subjects are with numbers 10, 22, 24, 15, 6, 1, 25, 11, 13, & 16.

| | | |
|---|---|---|
| 1. Robert | 11. Ken | 21. Steve |
| 2. Ralph | 12. Wilmer | 22. Larry |
| 3. John | 13. Alan | 23. Rick |
| 4. Andy | 14. Kevin | 24. Bruce |
| 5. Joel | 15. James | 25. Clyde |
| 6. Thomas | 16. Henry | 26. Sam |
| 7. Cooper | 17. Don | 27. Kent |
| 8. Maurice | 18. Walt | 28. Travis |
| 9. Terry | 19. Doug | 29. Woody |
| 10. Carl | 20. George | 30. Brian |

# Probability Sampling

## 1. Simple Random Sampling, cont.

Estimate hemoglobin levels in patients with sickle cell anemia.

1.  Determine sample size
2.  Obtain a list of all patients with sickle cell anemia in a hospital or clinic
3.  Patient is the sampling unit
4.  Use Lottery /table of random numbers method to select units from the sampling frame
5.  Measure hemoglobin in all patients
6.  Calculate mean and standard deviation of sample

## 2. Systematic Random Sampling

- Sampling interval (i/I)
- Population (N)
- Sample size (n)
I=N/n
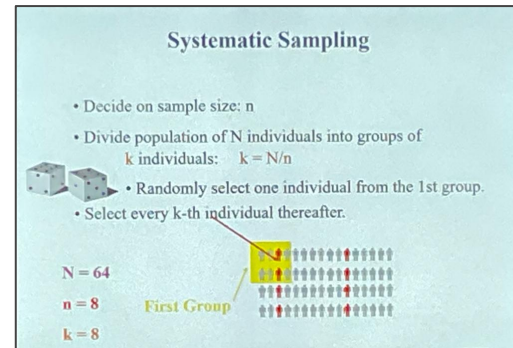
- **Technique:**
    - The sample is chosen by selecting a random starting point and then picking every i th element in succession from the sampling frame.
    - The **sampling interval i,** is determined by dividing the **population size N** by the **sample size n** and rounding to the nearest integer.[1]
    - Use "system" to select sample (e.g., every 5th item in alphabetized list or every 10th name in phone book)
- **Advantage:**
    - Sampling frame does not need to be defined in advanced
    - Easier to implement in the field
    - If there is unrecognized trends in the sample frame, the systematic sample ensure coverage of the spectrum of units
- **Disadvantages:**
    - Variance cannot be estimated unless assumptions are made

**First number** must be **random, if** it's **not** random then its called **systematic non-random sampling**

**Example:**
-   If a systematic sample of **500** students were to be carried out in a university with an enrolled population of **10,000**, the sampling interval would be:

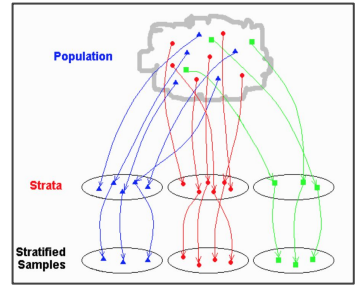$I = N/n = 10,000/500 = 20$

-   All students would be assigned sequential numbers. The starting point would be chosen by selecting a random number between 1 and 20. If this number was 9, then the 9th student on the list of students would be selected along with every following 20th student. The sample of students would be those corresponding to student numbers 9, 29, 49, 69...9929, 9949, 9969 and 9989.

[1] **For example,** there are 100,000 elements in the population and a sample of 1,000 is desired. In this case the sampling interval, i, is 100. A random number between 1 and 100 is selected. If, for example, this number is 23, the sample consists of elements 23, 123, 223, 323, 423, 523, and so on.
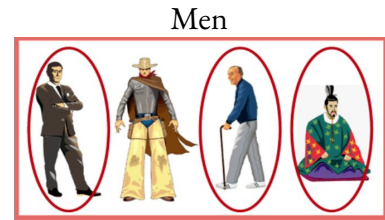
- **Technique:**
  - A two-step process in which the population is partitioned into subpopulations, or strata.
  - The strata should be mutually exclusive and collectively exhaustive in that every population element should be assigned to one and only one stratum and no population elements should be omitted.
  - Next, elements are selected from each stratum by a random procedure, usually simple random sampling.
  - The sampling frame comprises groups, or strata, with certain characteristics
  - A sample of units are selected from each group or stratum
- **Advantage:**
  - Assures that certain subgroups are represented in a sample
  - Allows investigator to estimate parameters in different strata
  - More precise estimates of the parameters because strata are more homogeneous.
  - Strata of interest can be sampled most intensively, e.g. groups with greatest variance
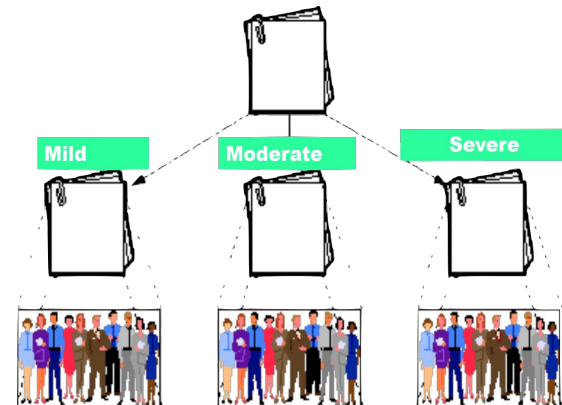  - Administrative advantages
- **Disadvantages:**
  - Loss of precision if small number of units is sampled from strata



Men



Women



Stratified Random selection for drug trial in hypertension:

| List of clients |
| **Strata** |
| Random subsamples of n/N |



## Sampling in Epidemiology

Stratified random sample, Assess dietary intake in adolescents:

1. Define three age groups: 11-13, 14-16, 17-19
2. Stratify age groups by sex
3. Obtain list of children in this age range from schools
4. Randomly select children from each of the 6 strata until sample size is obtained
5. Measure dietary intake

# 4. Cluster (Area) random sampling

**Technique:**

- The population is first divided into mutually exclusively groups of elements called clusters.
- Ideally, each cluster is a representative small-scale version of the population (i.e., heterogeneous group)
- A simple random sample of the clusters is then taken.
- All elements within each sampled (chosen) cluster form the sample.

**Advantages:**

- The entire sampling frame need not be enumerated ni advance, just the clusters once identified
- More economical in terms of resources than simple random sampling

**Disadvantages:**

- Loss of precision, i.e., weather variance, but can be accounted for with larger number of clusters

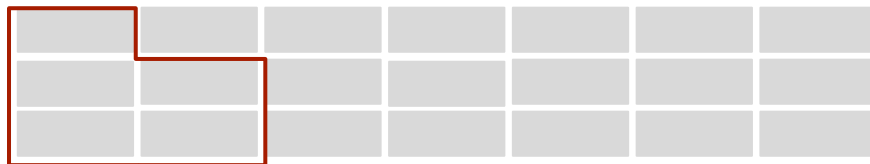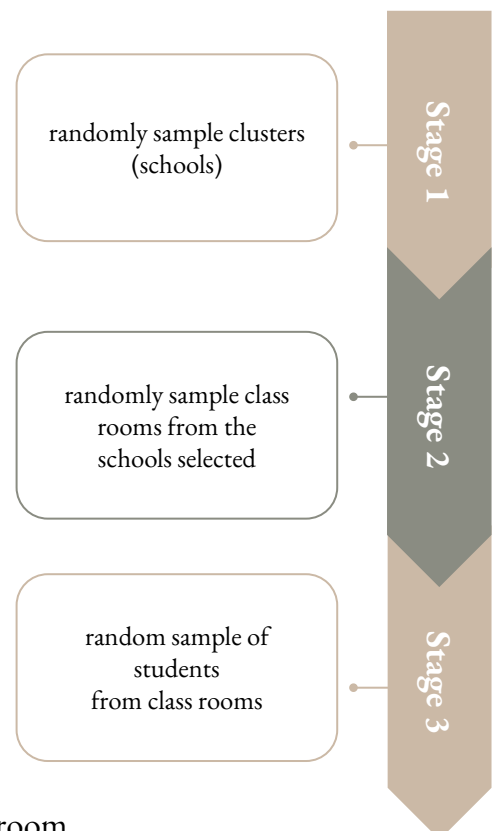E.g., Estimate the prevalence of dental carriers in school children:

1. Among the schools in the catchments area, list all of the classroom in each school
2. Take a simple random sample of classrooms or cluster of children
3. Examine all children in a cluster for dental carriers
4. Estimate prevalence of carriers within clusters, then combine an overall estimate, with variance
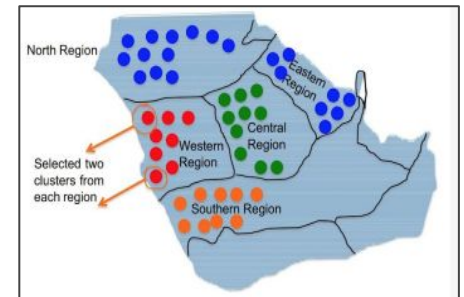


# 5. Multistage random sampling

- This sampling method is actually a combination of the basic sampling methods carried out in stages.

- Aim of subdividing the population into progressively smaller units by random sampling at each stage.



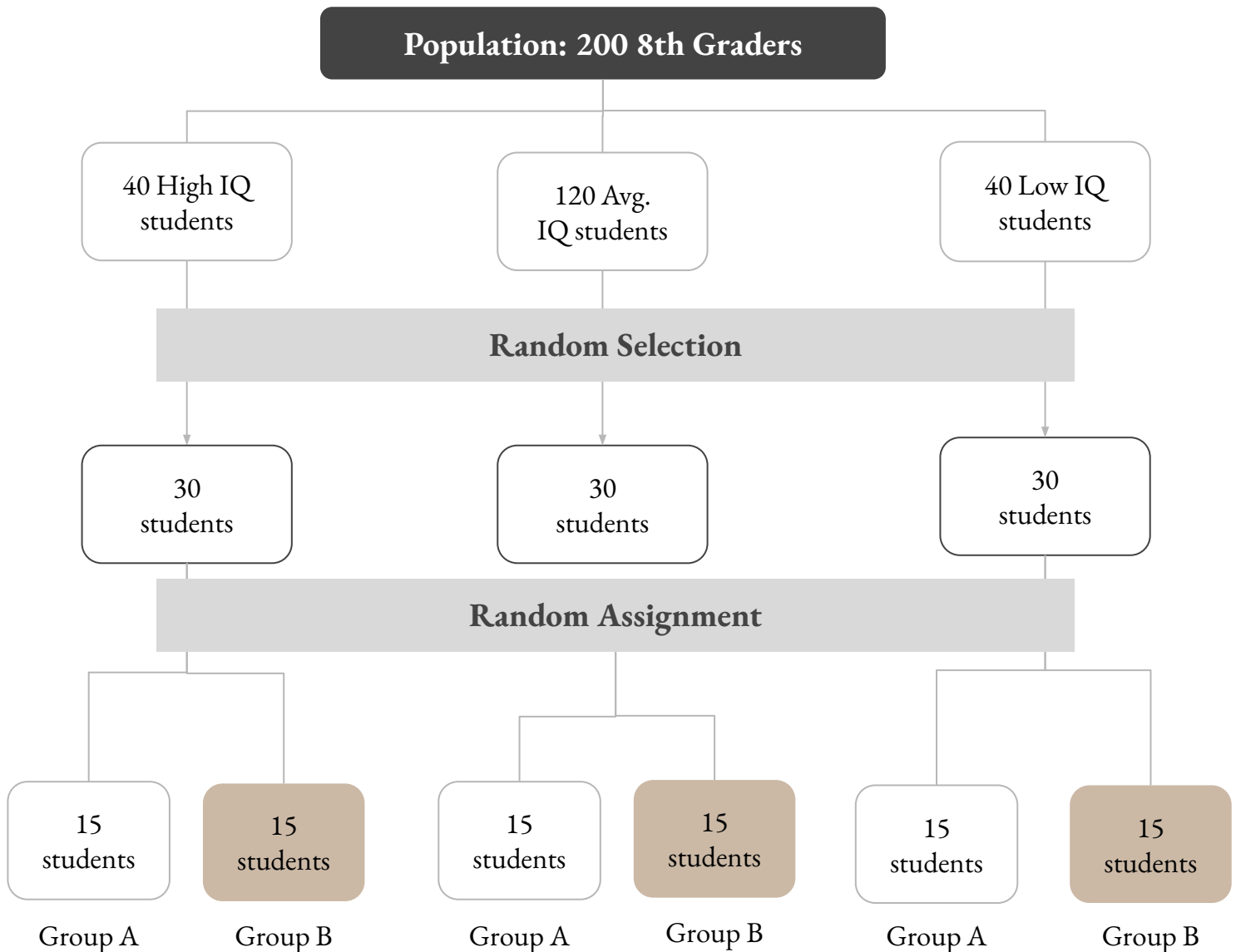**First stage:** randomly select x clusters (schools)

**Second stage:** within each school, randomly select y clusters (classrooms)

**Third stage:** randomly select x number of people from the classroom

**Extension of cluster sampling**

| | |
|---|---|
| randomly sample clusters (schools) | Stage 1 |
| randomly sample class rooms from the schools selected | Stage 2 |
| random sample of students from class rooms | Stage 3 |

# Random Selection vs. Random Assignment

| Random Selection | Random Assignment |
|---|---|
| • Every member **of the population** has an equal chance of being selected for the sample.<br><br>• Subject Selection (Random Selection): Choosing which potential subjects will actually participate in the study. | • Every member of the sample (however chosen) has an **equal chance** of being placed in the **experimental group or the control group.**<br>• Random assignment allows for individual differences among test participants to be averaged out. **Both the study group and control group must have similar characteristics otherwise you'll get bias.**<br>• Subject Assignment (Random Assignment): Deciding which group or condition each subject will be part of. |

**Population: 200 8th Graders**

| 40 High IQ students | 120 Avg. IQ students | 40 Low IQ students |
|---|---|---|

**Random Selection**

| 30 students | 30 students | 30 students |
|---|---|---|

**Random Assignment**

| 15 students | 15 students | 15 students | 15 students | 15 students | 15 students |
|---|---|---|---|---|---|
| Group A | Group B | Group A | Group B | Group A | Group B |

## 1. Convenience (Haphazard) Sampling

Convenience sampling attempts to obtain a sample of convenient elements. Often, respondents are selected because they happen to be in the right place at the right time.
- "Take them where you find them", nonrandom

Examples:
- Use of students, and members of social organizations
- Mall intercept interviews without qualifying the respondents
- Department stores using charge account lists
- "people on the street" interviews
- Distribution of survey via social media platforms
- Case series of patients with a particular condition at a certain hospital
- "Normal" graduate students walking down the hall are asked to donate blood for a study
- Children with febrile seizures reporting ot an emergency room.

Investigator decides who is enrolled in a study

- **Disadvantage**
  - Selection bias

**Page 4:** All non-probability sampling will be **used in qualitative** research **NOT quantitative** research.
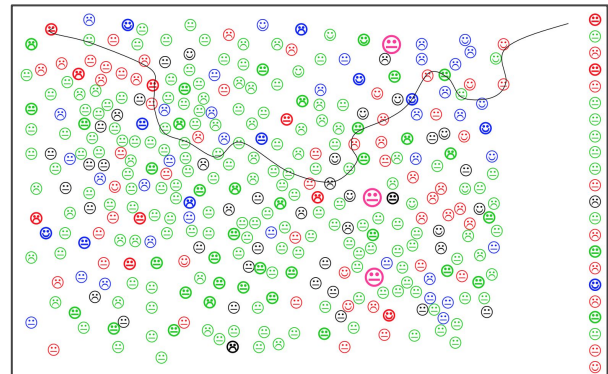


If you ask someone a sensitive question, for example do you use any illegal drugs? and he says no, while he is using them What type of bias is this?

Self-report bias

## 2. Judgmental Sampling (Purposive)

- Purposive sampling (**criterion-based sampling**)
  - Establish criteria necessary for being included in study and find sample to meet criteria.
- Solution: **Screening**
  - Obtain a sample of larger population and then those subjects that are not members of the desired population are **screened or filtered out.**
  - E.g., if you want to study smokers but can't identify all smokers.
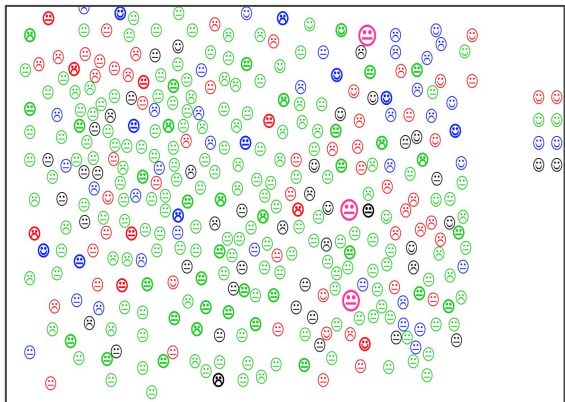
## 3. Deliberate (Quota) Sampling

Quota sampling may be viewed as two-stage restricted judgmental sampling.

- **Technique**
  - **The first stage** consist of developing control categories, or quotes, of population elements.
  - **In the second stage**, sample elements are selected based on convenience or judgment.
- **Disadvantage**
  - Selection bias
  - Cannot set quotas for all characteristics important to study
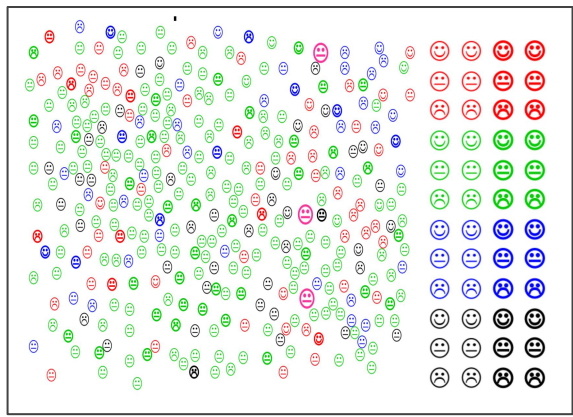


## 4. Snowball sampling

- In **snowball sampling,** an initial group of respondents are selected, usually at random.
- After being interviewed, these respondents are asked to identify others who belong to the target population of interest.
- Subsequent respondents are selected based on the referrals.
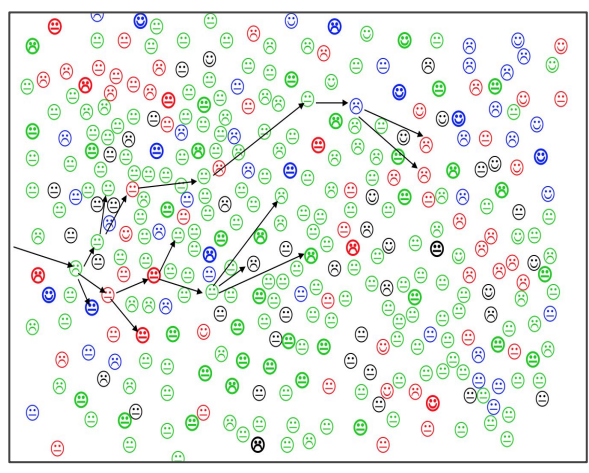


## 5. Consecutive sampling

- Consecutive series of consecutive patients with a condition of interest.

- Consecutive series means all patients with the condition within hospital or clinic, not just the patient that investigators happen to know about.

**Examples:**

- Outcome of 1000 consecutive patients presenting to the emergency room with chest pain.
- Natural history of all 125 patients with HIV associated TB during 5 year period.
- Explicit efforts must be made to identify and recruit **ALL** persons with the condition of interest.

## Conclusion

For any research, based on its study design and objectives an appropriate random sampling technique should be used.

**Prof. Shaffi: very important table**

| Probability sampling | Evaluation Criteria | Non-probability sampling |
|---|---|---|
| Conclusive | Nature of research | Exploratory |
| Larger sampling errors | Relative magnitude sampling vs non-sampling error | Larger non-sampling error |
| High [Heterogeneous] | Population variability | Low [Homogeneous] |
| Favorable | **Statistical Considerations** | Unfavorable |
| High | Sophistication Needed | Low |
| Relatively Longer | Time | Relatively shorter |
| High | Budget Needed | Low |

**Q1.** A survey will be given to 100 students randomly selected from the freshmen class at Lincoln High School. What is the population?

    A.    all freshmen at Lincoln High School
    B.    all students at Lincoln High School
    C.    The 100 selected students
    D.    All high school students in the US

**Q2.** A truck manufacturer selects 3 trucks at random from each of 6 models for safety testing. What type of sample is this?

    A.    Systematic random sample
    B.    Cluster random sample
    C.    Stratified random sample
    D.    Simple random sample

**Q3.** A quality control worker at a factory selects the first 10 items she sees as her sample for the day.What type of sample is this?

    A.    Convenience sample
    B.    Simple random sample
    C.    Systematic random sample
    D.    Stratified random sample

**Q4.** Inspectors for a hospital chain with multiple locations randomly select some of their locations for a cleanliness check of their operating rooms. The inspectors check every operating room in the hospitals that were chosen?

    A.    Systematic random sample
    B.    Cluster random sample
    C.    Convenience sample
    D.    Simple random sample

**Q5.** Security workers at an airport randomly choose one of the first 505050 people to pass through a checkpoint for extra security screening. After that person, they choose every 50$^{th}$ person who passes through for extra screening as well.?

    A.    Systematic random sample
    B.    Cluster random sample
    C.    Convenience sample
    D.    Simple random sample

**Q6.** Each student at a school has a student identification number. Counselors have a computer generate 50 random identification numbers, and the students associated with those numbers are asked to take a survey?

    A.    Systematic random sample
    B.    Cluster random sample
    C.    Convenience sample
    D.    Simple random sample

Answers:
1: A | 2: C | 3: A | 4: B | 5: A | 6: D

القادة :

نواف التركي

ريان الغنامي

عبدالله الشهري

وهى المتحمي

الأعضاء:

| | |
|---|---|
| عبدالله التركي | عبدالله المياح |
| محمد الزير | عبدالله النجرس |
| عثمان الدريهم | تركي العتيبي |
| عبدالعزيز القحطاني | عبدالله القرني |
| ناصر الغيث | عامر الغامدي |
| سعد السهلي | سعد الاحمري |
| رائد الماضي | معاذ آل صلام |
| سعود الشعلان | محمد الحصيني |

رغد النظيف

ديما الجريبة

شهد البخاري

نوف الضلعان

أثير الاحمري

وعد ابو نخاع

ثراء الهويش

في الدوسري

منار الزهراني